

TRIBHUVAN UNIVERSITY

Institute of Science of Technology

2066

Bachelor Level/ First Year/ First Semester/ Science

Computer Science and Information Technology (Stat. 103)

(Probability and Statistics)

Full Marks: 60

Pass Marks: 24

Time: 3 hours.

Candidates are required to give their answers in their own words as far as practicable.

All notations have the usual meanings.

Group A

Attempt any Two:

(2x10=20)

1. Define the following three measures of locations – mean, median and mode – and clearly state their properties. Write down a situation where mode is preferred to mean. Score obtained by 14 students in a test are given below. Compute mean, median and mode.

42	39	45	55	38	35	60	55	55	65	40	43	35	37
----	----	----	----	----	----	----	----	----	----	----	----	----	----

2. Explain the terms – sample space and events of a random experiment. State the classical and the statistical definition of probability. Which of the two definitions is the most useful in statistics and why? A survey of 300 families was conducted to study income level versus brand preference. The data are summarized below.

Income level	Brand			
	Brand 1	Brand 2	Brand 3	Total
High	55	45	20	120
Medium	45	25	25	95
Low	25	35	25	85
Total	125	105	70	300

If a family is selected at random, then compute the probability that (a) the family belongs to high income group, (b) the family prefers Brand 3, and (c) the family belongs to the low income group and prefers Brand 3.

3. Make a clear distinction between correlation coefficient and slope regression coefficient. A school teacher believes that there is a linear relationship between the verbal test score (Y) for eighth graders and the number of library books checked out (X). Following are the data collected on 10 students.

X	12	15	3	7	10	5	22	9	13	7
Y	77	85	48	59	75	41	94	65	79	70

The above data reveal the following statistics:

$$\Sigma X = 103, \Sigma Y = 693, \Sigma X^2 = 1335, \Sigma Y^2 = 50447, \Sigma XY = 7881$$

- (a) Compute the correlation coefficient r between X and Y . Interpret the meaning of r^2 .
- (b) Fit a simple linear regression model of Y on X using the least square method. Interpret the estimated slope regression coefficient.

Group B

Attempt any eight questions:

(8x5=40)

4. State with suitable examples the role played by the computer technology in applied statistics and also the role of statistics in Information Technology.
5. Define discrete and continuous random variables with suitable examples. A continuous random variable X has the following density function.

$$f(x) = \begin{cases} kx(1-x) & \text{for } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Find the value of k show that the total probability would be 1. Also find $E(X)$.

6. Assume that the two continuous random variables X and Y have the following density function

$$f(x) = \begin{cases} \frac{6-x-y}{8} & \text{for } 0 < x < 2, 2 < y < 4 \\ 0 & \text{elsewhere} \end{cases}$$

Find (a) marginal density function of X and (b) conditional probability $P(2 < y < 3 | x=1)$.

7. In a binomial distribution with parameters n and p , prove that mean and variance in binomial distribution are correspondingly np and npq , where $q = 1 - p$.
8. The systolic blood pressure of 18 years old women (X) is normally distributed with a mean of 120 mm Hg and a standard deviation of 12 mm Hg randomly selected 18 years old women. Compute the following probabilities:
(a) $P(X > 150)$ (b) $P(X < 115)$ (c) $P(110 < X < 130)$
9. If $X_1, X_2, \dots, X_n$ are n independent random variables each is distributed as normal with mean μ and variance σ^2 , then derive the distribution of $\sum_{i=1}^n X_i$.
10. Write the density function of negative exponential distribution, and derive its mean and variance.
11. Obtain the maximum likelihood function of n independent random sample drawn from a normal population with unknown mean μ and unknown variance σ^2 , and, using the principle of maximum likelihood method of estimation derive the estimators of μ and σ^2 .

12. A survey of 100 percents of first and second grade children revealed that the number of hours per week their children watch television (X) had an average of 25.8 hours and standard deviation of 4.0 hours. The problem is to determine whether there is statistical evidence to conclude that μ (population mean of X) exceeds 25 hours. Set up appropriate null and alternative hypothesis and carry out appropriate test at 5% level of significance.
13. A standardized psychology exam has a mean of 70. A research psychologist wished to see whether a particular drug had an effect on performance on the exam. He administered exam to 18 volunteers who had taken the drug, and obtained the following scores: 68, 71, 71, 65, 64, 70, 70, 64, 71, 73, 62, 78, 70, 69, 76, 67, 69, 72, which yielded $\bar{x} = 69.4444$ and $s^2 = 16.8497$. The problem is to determine whether there is statistical evidence suggesting that taking drug reduces one's score on the exam. Set up appropriate null and alternative hypothesis and carry out the test at 5% level.